

WikiMining, WikiNetworks and Wikinomics

Laurence Lock Lee, Optimice (llocklee@optimice.com.au)

James Matheson, CustomWare (james.matheson@customware.com)

Stewart Mader, Grow Your Wiki (slmader@gmail.com)

Introduction

While Wikipedia is responsible for introducing the Wiki to the masses, it was Tapscott and Williams (2006) who built the linkage to business and economic benefit through their influential book on “Wikinomics”¹. This paper reports on the results of a research project conducted to explore the value achieved through participation in a wiki community. The project developed WikiMining techniques² to infer social networks that may exist within a wiki community. The Wikipatterns.com community launched in 2007 to develop a community around accelerating wiki adoption, was used as the vehicle for the research. The overall aim of the project was to take Wikinomics to a more granular level where specific benefits can be associated with specific social networking patterns. Ultimately it is anticipated that the identification of particular collaborative patterns through the use of Net Mining techniques, will provide an early indicator of economic success or otherwise.

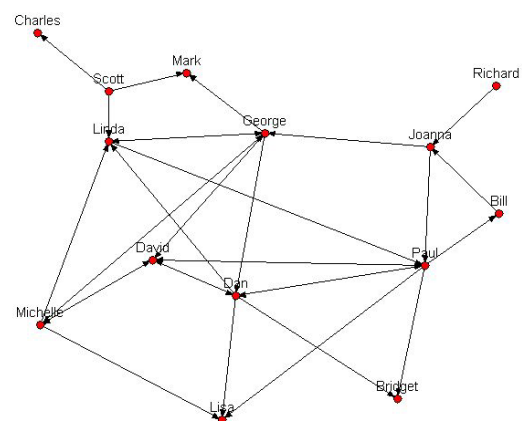
¹ Tapscott D. and Williams A. (2006), “*Wikinomics: How Mass Collaboration Changes Everything*”, Portfolio.

² See http://www2.optimice.com.au/Index_of_Industry_Maps.php for examples of Net Mining

The paper will provide some background on Social Network Analysis (SNA) techniques and the use of social network mining to discover social networks. This will be followed by a description of the methods used for the study, the results achieved and finally a discussion on the implications of the findings.

Net Mining and Social Network Analysis

SNA is a well established technique³ for the study of social networks. It is most associated with the development of the sociogram, a graphical representation of connections between individual actors:



³ See www.insna.org for a comprehensive description



The ‘actors’ are typically individuals, but can also be companies or organizations, depending on the level of analysis being conducted. The links denote relationships, which can be identified with friendship, professional connections, advice or knowledge sharing and the like. The relationships can be directional e.g. Bill nominates Joanna as a connection, which may or may not be reciprocated. The typical network is called an “ego network” where the network is identified by direct nomination of connections, usually through the use of a social network survey. A second form of network is called an “affinity network”. These networks are inferred through common attributes of the actors. For example, a link between Dan and Paul might be inferred if they belong to the same sports club, or sit on the same board of directors. The strength of connection or “ties” can also be inferred in an ego network by actors nominating a relative strength of tie in the survey, or for an affinity network, by the number of common attributes of the actors. From the sociogram a plethora of social network metrics have been developed to identify key players, network cliques and clusters and the like.

While the social network survey has been the staple means for collecting social network data, it suffers from the same problems of all survey techniques in that they can only be repeated at infrequent intervals and can only achieve a subset sample of responses from the whole of the community under study. With the advent of electronic collaboration tools, starting with e-mail, and moving on to discussion lists and now web 2.0 tools, the potential now exists to infer social networks through

analyzing the logs of the collaborative software in use by the community under study. Referred to as ‘Net Mining’⁴, the technique allows the whole community to be surveyed at whatever frequency is desired. The technique analyses actual collaborative actions, so can arguably be a more accurate reflection of networking patterns than the reflective responses provided by survey respondents. The downside is that it is still too early to tell if the patterns identified by Net Mining techniques are an accurate reflection of the true social constructs being inferred. This will undoubtedly be addressed in the coming years as the technique matures, with this project hopefully contributing to the effort.

What we did

As indicate earlier, the Wikipatterns.com community was used as the vehicle for this research. The Wikipatterns.com site was established by Stewart Mader at the beginning of 2007 to build a community around practitioners looking to accelerate the adoption of wikis within their respective organizations. A key theme of the site is the development of particular adoption archetypes, reflecting good and bad adoption patterns. As of June 2008, there were some 200 registered members. The readership of the site is likely to be substantially more than this, though unfortunately data on non-registered readers was not available for this project.

⁴ See MIT center for collective intelligence for examples <http://www.ickn.org/index.html> or this paper provides an example of an early application using a discussion list http://www.optimice.com.au/upload/Does_your_Community_Leave_a_Digital_Footprint.pdf



A method of net mining was designed to extract network data from the usage logs from the wiki. Essentially the technique inferred a connection between registrants from their co-editing or commenting on particular wiki pages. The data was then fed into a SNA tool to create an affinity network of contributors to the wiki. For simplicity, a relationship between editors was inferred if they edited the same page in the wiki.

To identify the value being achieved through participation in the wiki community, an on-line survey⁵ was conducted of the registered members. The survey initially asked the respondents to rate the value that they had achieved from being a member of the community. It then asked them to nominate the type of role that they had adopted in the community. The roles were defined as:

- Core member – foundation member
- Commentator – leaves comments on pages
- Sub-editor – edits pages for clarity without changing the content
- Collaborator – edits pages
- Browser – reads
- Subscriber – registers with the wiki
- Explorer – posts a page but does not collaborate further unless it engages with others

These archetypes were identified as typical of those found of a wiki participant as well as those that could potentially be inferred from the net mining. The survey then went on to ask each respondent to nominate from the

list of other registrants, those that they knew or had interacted with. They were asked to classify these connections by whether they were pre-existing i.e. existed before participation in the wiki or if they were created from participation in the wiki. They were also asked to nominate the relative value of these relationships to them. The ego networks developed from the survey were then matched against the affinity networks inferred from the net mining, looking for common social network metrics. This was limited to network centrality (number of connections for each node), as this measure has been shown to be robust for sample sizes smaller than the full community⁶. The particular centrality measure used was the ‘degree’ centrality measure which simply counts the number of connections to other actors. The measures were normalized to cater for the different network sizes.

The archetypes identified from the survey were used to identify links with the nominated value received by the participants. The nominated value of new relationships created from participation in the wiki community was also used as a measure of value received.

What we found

A network map of the raw editor / page interaction is shown below:

⁶ Borgatti, S. P., Carley, K., & Krackhardt, D. (2006). Robustness of centrality measures under conditions of imperfect data. *Social Networks*, 28, 124-136.

⁵ The survey tool used at www.onasurveys.com was designed specifically for SNA use.

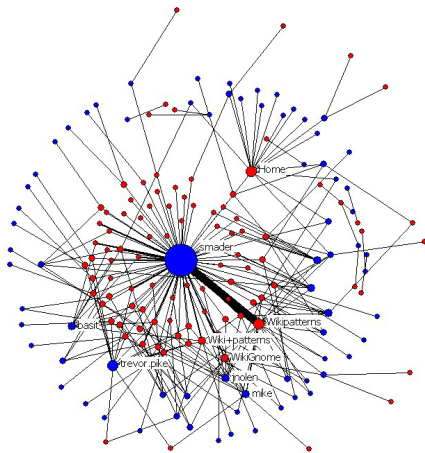


Figure 1 – Wiki Editor – Page Associations

The blue nodes identify page editors and/or creators and the red nodes are pages. The links identify an editing activity, with the thickness of the lines identifying the relative number of edits made to the page by a particular editor. The size of the nodes reflects the relative number of edits made by a given editor or edits made to a given page. As one can see, the Wikipatterns.com founder, ‘smader’ has been centrally active in

generating and maintaining pages, as one would expect from a founding member. It is evident however, that he is now being joined by a growing number of editing colleagues.

Relating the ‘Mined’ to ‘Surveyed’ Networks

From the above data the network map of connections between editors was developed. This map essentially infers a relationship between editors on the basis of a shared interest in a given page. Again the strength of the relationships can be inferred by the number of pages a pair of editors has a common interest in. This relationship map can then be compared with the relationship map identified through the survey, enabling a comparison between the ‘inferred’ relationship map as mined from the wiki activity, and the nominated relationship map as identified through the direct surveying of editors.

“Mined (inferred)” Relationships

Stated Relationships (from survey)

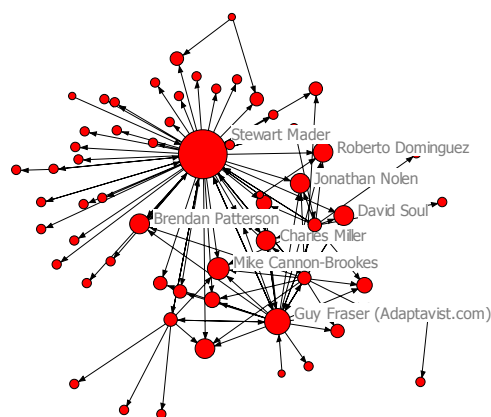
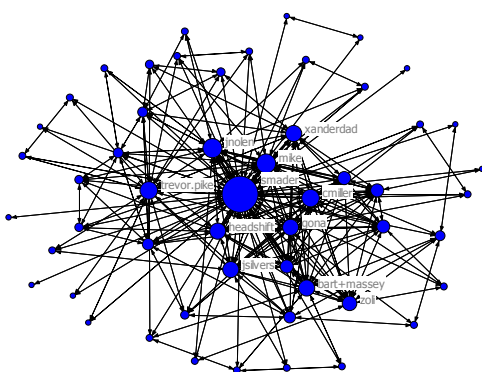


Figure 2 – Comparison of ‘Mined’ and ‘Stated’ Relationship Networks

The appearance of the two maps differs to some extent as for the surveyed relationships the map only contains responses from 40 of the around 200 registered participants, who nominated some 70 contacts, whereas all active editors are included in the 'mined' map.

What is important however is how well the inferred relationships mimic the stated relationships. To do this the most centrally connected nodes in both maps were compared with the following results:

"Mined (inferred)" Relationships		Stated Relationships (from survey)	
name	Degree	name	Degree
Stewart Mader	94.74	Stewart Mader	57.14
Charles Miller	26.32	Guy Fraser	28.57
Trevor Pike	22.81	Charles Miller	7.80
Mike Cannon-Brooks	22.22	Mike Cannon-Brookes	6.50
Jon Nolen	21.05	Jon Nolen	5.20

Figure 3 – Top five editors ranked by Degree Centrality

One can see from the ranking table that four of the five most central editors are common to both the mined and surveyed networks. The lower values for the surveyed results are due to many of the nominated contacts not being survey respondents, resulting in lower scores on average. A more comprehensive correlation of the full list of editors as identified by the survey respondents and their centrality in the mined network shows a respectable and statistically significant 0.77 correlation.

From these results we can cautiously suggest that the relationship networks as identified through mining of the wiki logs are representative of the directly stated relationships as identified through survey means. This is an important result for the viability for using wiki mining to both infer actual relationships between

wiki participants and then to predict performance based on those patterns.

Wikipatterns.com Performance Results

An important objective of the survey of Wikipatterns.com participants was to identify value achieved from participation in the wiki. A key performance measure was identified as the personal value that participants felt they achieved from their participation. Of the 37 valid responses to this question 15 (40%) of the respondents identified a value score of 4 or 5 on a 5 point scale. When this sub sample of respondents are identified with the archetype roles they nominate that they played we get the following result:

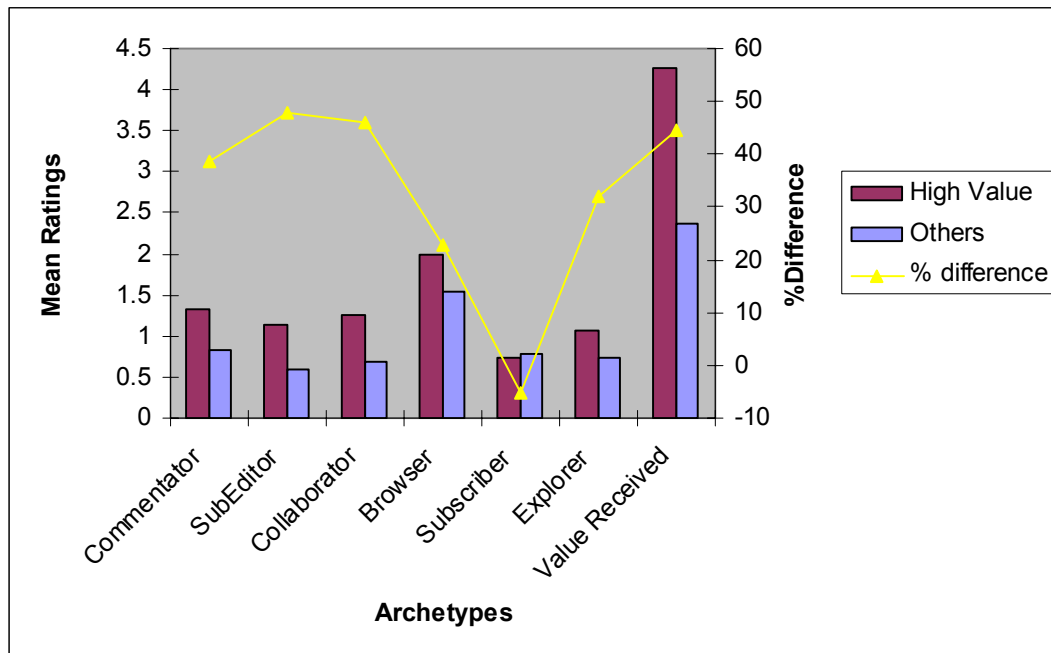


Figure 4 – Archetype roles differentiating those who gained high value from wiki use to the remainder

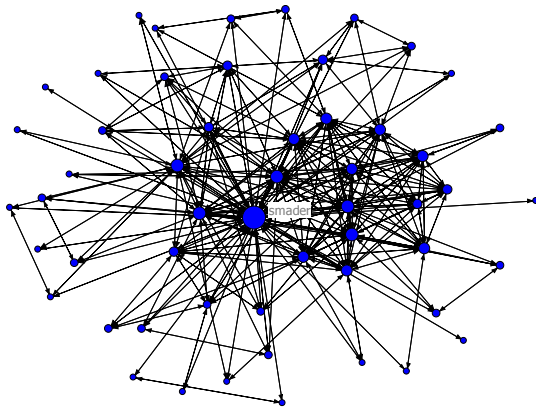
The ratings relate to the mean score that this sample of respondents gave to their relative conduct of each archetype role on a 5 point scale. The sub sample labeled as receiving ‘high value’ from participation in the wiki scored ‘value gained’ as 4 or 5 on a 5-point scale. This group was then compared with the remainder of the respondents. We can see from these results that those respondents who most valued membership were more active in each of the archetype roles with the exception of ‘subscriber’, as each respondent had to be subscribed to receive a survey. Looking at the percentage differentials, we can see that those that took on the more active roles of commenting, collaborating, editing and to a lesser extent, exploring, were associated with gaining more value from their participation. In other words one could assume that those more active

participants also gained more value from being members.

The next area of analysis looked at the health of the community as a whole. Research on communities of practice⁷ identify stages that communities pass through from launch, to growth and then sustainability. After some 18 months of operations the wikipatterns community may be preparing to transition from a growth to a sustaining stage now. A good test of the sustainability of a community is the emergence of many potential leaders, enabling the leadership responsibilities to be rotated or shared over time. The following graphic identifies the community with and without its founder, Stewart Mader:

⁷ For example, Wenger, McDermott and Snyder, (2002) *“Cultivating Communities of Practice”*, Harvard Business School Press.

Founder as central connector



Founder not present

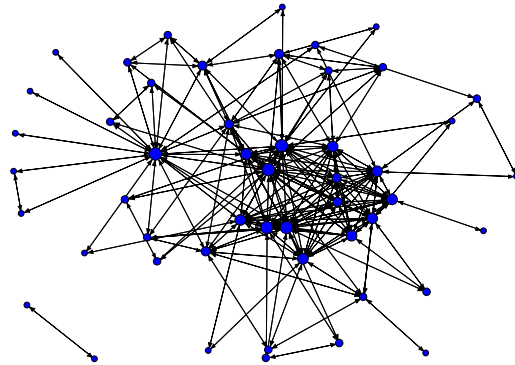


Figure 5 – Community sustainability assessment

One can see that by taking out the founder and his connections from the network, the network does appear less dense. Importantly however, there exist several nodes that are sufficiently connected in their own right, to ensure that the community can be sustained even if the founder were to stop participating. This is good news for the community and shows a maturity that is required to ensure a good level of sustainability should a few critical members cease to participate.

The final area of analysis looked at how new social relationships were being formed through participation in Wikipatterns.com. In the survey respondents were asked to nominate contacts and their perceived value of the relationship with these contacts. They were also asked to nominate the type of relationship as: a prior relationship > 2 years; a prior relationship < 2 year; a relationship created through the wiki but now extending beyond it; and a relationship through the wiki only. The

number after the classifications in the key indicate the number of nominated relationships in this class e.g. 16 of the nominated relationships had pre-existed for more than 2 years. Each bar in the bar chart reports the results against the relationship value nominated by the respondents e.g. for those respondents who nominated the relationships as ‘very important’, some 35% were pre-existing relationships of more than 2 years, a further 25% were pre-existing relationships of less than 2 years and interestingly some 35% were new relationships generated from meeting in Wikipatterns.com. As one can see from the chart, of the 22 new relationships developed through meeting in Wikipatterns.com, the majority have been rated as ‘very important’ by the respondents. This is an extremely positive result for the community in its ability to broker highly valued new relationships through participation in the wiki.

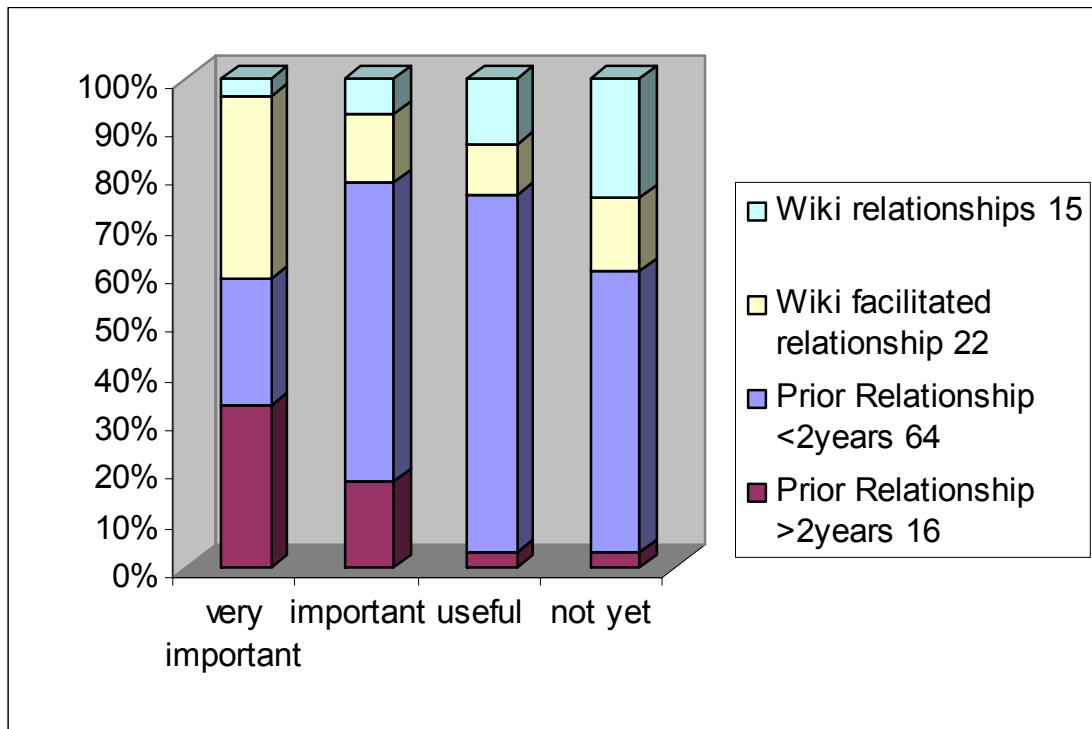


Figure 6 – Brokering Valued Relationships via the Wiki

Overall the results identify several important benefits that have been generated through the Wikipatterns.com community. At the time of the research there were just over 200 registered members. The combination of wiki mining and direct surveying has shown that members who are active participants i.e. collaborators, editors, commentators or page creators are gaining the most value from their membership. The network analyses shows that new leaders are emerging and that the community could sustain itself should the founder cease to participate. Finally the results show that the community has been successful in generating many new and highly valued relationships that now span beyond the Wikipatterns.com site.

Summary and future directions

This research set out to identify how the value that can accrue from participation in a wiki is related to the collaboration patterns of the members. Social network analysis techniques were used to help visualize the collaboration patterns and a survey conducted to identify the value achieved by the members. These SNA patterns were generated both from the mining of the activity logs of the wiki and directly surveying the members for their contacts. An important finding in this exercise was the mined network maps which inferred relationships through collaboration patterns correlated well with the patterns that were directly nominated through the survey. That is, those identified as centrally connected



actors in the network inferred from the network mining corresponded well with those nominated directly through the survey. The correspondence however was not perfect and there were instances where central connected actors identified through network mining were not in the direct nomination list and visa versa. This suggests that not all socially centrally connected people are active wiki collaborators and not all wiki collaborators are socially central to the network. Why this is so is an opportunity for follow on research. The correspondence however was encouraging enough to promote greater use of the network mining approach for relating collaboration patterns to performance, as a more efficient means for collecting such data than traditional survey means. The major advantage of the use of network mining is that all logged activity can be used over any time frame without the need for the intrusion on the membership that a survey requires.

In terms of the Wikipatterns.com community itself, the results identified that a majority of members were gaining high value from their membership, especially those more actively engaged. The results also showed that the density and breadth of collaboration suggests the

community will be self sustaining into the future.

In terms of future research, there is still much more to do in directly linking collaboration facilitated through wikis to tangible business benefits. In this light more case studies will be analysed to identify situations where wikis are being used to support project situations, where performance metrics of time, cost and benefits can be measured independent of the collaborative team delivering the projects.

While it is still early days in the adoption of Enterprise 2.0 tools, the ability to identify the best mode of use of such tools and then to relate this to business value achieved will be a critical requirement for the successful adoption of wikis and other Enterprise 2.0 tools by mainstream business.

Acknowledgements

We would like to acknowledge the participation of the Wikipatterns.com members who participated in our survey and hope that these results will be of assistance to you in your endeavors in gaining business support for your wiki implementations.

About Optimice (www.optimice.com.au)

Optimice's objective is to help our clients optimise their internal and external business relationships. Optimice has extensive experience in researching, analysing and facilitating improved business relationships. In particular, Optimice aims to assist its clients successfully navigate their way through the growing complexity of alliance networks, multiple sourcing contracts, off-shore relationships, organisational restructures and the like, as well as help their staff to become more astute networkers.

About CustomWare (www.customware.net)

CustomWare is a global technology services company specialising in connecting systems (integration) and people (collaboration). The CustomWare team designs and integrates new and existing business processes and technologies. CustomWare has helped over 100 clients globally deploy and maximize their use of wiki technologies since 2002.

About Grow Your Wiki (<http://www.ikiw.org/services>)

Grow Your Wiki was founded by Stewart Mader in October, 2005 and offers insight, examples, case studies, and advice on enterprise wiki adoption and usage inside organizations. The key to successful wiki adoption is to create the conditions for steady growth over a period of time. This gives teams the time and attention they need to be successful, and allows the people managing the wiki to ensure the service grows appropriately to meet the needs of the expanding user base. As people experience the benefits of wiki use, they will spread buzz, and help build steady demand. Stewart Mader provides consulting services on a retained or subscription basis.